DOCUMENT RESUME

ED 347 706 EC 221 067

AUTHOR Shimizu, Naoji
TITLE Application of Single Subject Experimental Designs and Its Issues.
PUB DATE Feb 89
NOTE 12p.; Article includes an English abstract. RIEEC Report is published by the Research Institute for the Education of Exceptional Children.
PUB TYPE Journal Articles (080)
LANGUAGE Japanese
JOURNAL CIT RIEEC Report; v38 p65-74 Feb 1989

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Disabilities; Educational Research; Psychological Studies; Research Design; *Research Methodology; Research Problems; Statistical Analysis
IDENTIFIERS *Single Subject Research Design

ABSTRACT

In psychological and special education research, group comparison experimental designs are more often used than single subject experimental designs. However, the application of group designs poses many difficulties, such as the collection of large numbers of homogeneous subjects for a particular experimental condition, the averaging of individual subjects, and the use of statistical significance tests. In the development of behavior analysis since the 1960's, new experimental design alternatives have appeared which emphasize direct measurement of treatment effects and causal analysis between independent variables and dependent variables. This paper describes single subject experimental designs and discusses general issues in their use. (JDD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

APPLICATION OF SINGLE SUBJECT EXPERIMENTAL DESIGNS AND ITS ISSUES

Naoji SHIMIZU
Division for the Education of the Mentally Retarded

ED347706

EC 231067

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Kietsu Fujiwara
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

BEST COPY AVAILABLE

特殊教育研究施設報告
1989, 38, 65-74

単一被験者実験計画法の利用とその諸問題

清水直治*

キーワード：単一被験者実験計画法　集団比較実験計画法　統計的有意性検定
処遇効果　信頼性　妥当性

はじめに

単一被験者実験計画法 (single subject experimental design) あるいは1事例実験計画法 (single case experimental design) とは，実験条件の作用や処遇の効果を判定するために，1人の被験者を対象に，変数操作や治療処置を反復して実施する形式の実験計画法である。それによって，操作した要因（独立変数）と標的とされた行動（従属変数）の変容との因果関係の分析が可能となり，その標的行動を変容させるための制御変数の特定ができることになる。

現在使われているような単一被験者実験計画法は，徹底して個別事例の分析を重視した Skinner (Skinner, 1953) を始祖とする実験行動分析や応用行動分析において，個体の行動変化を検証する方略として1960年代以降発展してきたものである (Sidman, 1960 ; Baer, Wolf, & Risley, 1968 ; Kratochwill, 1978 ; Barlow & Hersen, 1984) が，それ以前の，心理学を厳密な科学として位置づけようとする研究の経緯をみると，この単一の個人を被験者とする実験計画法はむしろ中心的に用いられてきたのである。例えば，Wundt は最初に実験的方法によって個人の経験を研究し，Ebbinghaus は，記憶実験用の無意味綴りを自分自身に課し，記憶と忘却に及ぼす要因を見出した。また Pablov もその条件反射の現象を，1個体で発見した後に，複数の個体に応用しその現象の事実を確認している。

他方，臨床心理学や精神医学の分野でも，精神分析や心理療法においてみられるように，事例ごとの詳細な検討が主要な研究方法として使われてきた。しかしながらこの分野での個人の扱い方は，先に述べたような実験的方法ははとんど用いられず，その大多数は，治療や処遇に影響を及ぼす要因を統制する手続きをまったく含まない，単なる治療経過の報告というべきものである。そこでは，治療者の臨床的印象が述べられ，治療の効果が主観的に記述されることになる。したがって，たとえ患者の主訴が消失したとしても，それに関連する要因が特定できず，得られた結果の一般化が必ずしもできなくなってしまう。

こうして臨床の分野においても，1940年代になると，科学的な方法に則って，治療効果を判定すべきだとする考え方が強くなってきたが，その方法とは，すでに実験心理学の研究で多用されていた集団比較実験計画法 (group comparison experimental design) である。治療を受けた群と治療を受けない群（統制群）とを比較することによって，そこでの治療介入の効果の検討が試みられた。

Fisher の推測統計学 (Fisher, 1925) を基礎にしたこの集団比較実験計画法は，1930年代に開発され，それ以降，むしろ単一の被験者を対象とする実験研究にとって代わって，心理学あるいは特殊教育学の分野でも盛んに利用されるようになってきた。その趨勢は今日においても衰えを知らない。ところがこの集団比較実験計画法も，その限界や制約を熟知して適用しなければ，多くの誤解や弊害をもたらすことになる。

本稿では，集団比較実験計画法との対比において，単一被験者実験計画法の適切な利用のために，その特徴とそれが含み持ついくつかの問題点について考察することを目的とする。

1. 集団比較実験計画法と統計的推測

複数の被験者を実験条件に即していくつかの群に分け，それぞれの群にたいする実験操作の結果について，推測統計学の手法を使って，その群間の差異を統計的に検定することによって，被験者の属する母集団を推定するという集団比較実験計画法は，心理学や特殊教育学に関する実験や調査において，現状では，非常に多く利用されている。

しかしながら，集団比較実験や調査の結果に統計的検

* 精神薄弱児教育研究部門

定を実施し，そこで有意差が得られたからといって，その研究の結果が科学的に支持されたというわけではない。実験手続きや結果の解釈に関して，集団比較実験計画法は以下に述べるような，固有の制約や問題を内在させていることが指摘されている(Morrison & Henkel, 1970；橘，1986)。

(1) 標本の単純無作為抽出の困難さ

統計的推測による有意性検定の意味するところは，推測統計学のさまざまな手法を用いて，得られた標本（ここでは被験者）からその標本が属する母集団の性質を推論することである。つまり，被験者それ自身ではなく，つねに母集団を問題としており，その被験者の所属した母集団を推定することがその目的なのである。

集団間の有意性検定を実施するには，その前提として次の3つの条件が，それぞれの母集団において満たされている必要がある。すなわち，①正規性 …… 標本を抽出した母集団は，その変数に関していずれも正規分布をする，②等分散性 …… それぞれの母集団の分散は等しい，③単純無作為標本抽出 …… 標本は各々の母集団から無作為に抽出されている，ということである。この標本の無作為性とは，母集団のすべての要素のなかから乱数表によって標本を選び出すという抽出手続きをさし，それによってまた，観測値の独立性が保証されているということが，必要な要件である。

①および②の条件が満たされないときには，通常は，χ^2 検定などのノンパラメトリック検定を行うことになる。ところが③については，多くの研究が無作為な標本の抽出という前提を満たしていないばかりか，実際にはその手続きをとることが非常に困難であることが，集団比較実験計画法の最も大きな弱点となっている。被験者が無作為に選び出されていないのであれば，統計的な推測にもとづいて母集団を推定しようとすることは誤りであり，ましてや，実験結果の一般化が非常にあいまいなものとなってしまう。

事実，特殊教育の分野の実験の中には，無作為な抽出によらない被験者からなる各集団の有意差を計算し，その結果を過剰に一般化してしまう記述が散見される。例えば，その母集団が不明確であり，しかも無作為抽出もされていない1群10名程度の「自閉症」についての実験の結果を，すべての「自閉症」に一般化できるように記述することは誤りである。最も確実なのは，そこで選び出した被験者全員を母集団として，結果を限定的に述べることである。その際には，母集団の全要素を入手していることになり，推測統計の適用は不必要となることも自明であろう。

(2) 集団間の変数のマッチングの誤用

集団比較実験計画法では，各々の群の条件を均質にするという理由から，しばしばMAやIQの等しい被験者どうしを対にしてから，それぞれの群に配置するということが行われる。これを変数のマッチングというが，実際はこのような群への割り付けは，無作為に実施されるわけではなく，また標本の無作為性が低くなり独立性がなくなってしまう。したがってそこに統計的検定を適用しても，もたらされる数値は，事実を正確に反映していないことになる。

(3) p値による有意性の基準

統計的有意性検定では，要するに，帰無仮説(H_0)が棄却できるか，あるいはできないかを判断しようとしているのであり，一般に有意性の基準として採用されている$p<0.05$は，H_0が正しいとすると，得られた標本の差がもたらされる確率は5%より小さいということである。しかしそれは同時に対立仮説(H_1)が95%正しいということを示すものではない。統計的に有意差がもたらされたということは，H_0が正しいとした場合に，観測された結果が得られる確率について述べているにすぎない。p値が大きいことは，決して実験条件や治療の効果の大きさを示すものではないのである。そして有意性検定の結果は，実験者の設定した仮説の妥当性の指標となるものでもない。またp値は$p=0.05$を境に有意性の有無が判断されるため，例えば$p>0.05$では有意ではないということで，その研究そのものの価値にも影響が及んでしまうという不都合さもある。

(4) 個人への結果の適用の不適切さ

集団比較実験計画法では，実験の結果は集団の平均値や分散で示され，個人の変動は誤差として処理される。この平均値や分散は集団の構成員の一人ひとりを反映したものではなく，結局は，そこで得られた情報をそのまま個別に適用しようとすることは，適切ではないことがわかる。

以上のような問題点とともに，集団比較実験計画法の利用においては，つねに多数の被験者を集めることに付きまとう問題が存在する。そこで，問題行動や行動変容のメカニズム，あるいはそれらに直接影響を及ぼす要因を見出そうとする研究者や臨床家は，1人の被験者や患者を扱うことへと関心を向けていった。

II. 単一被験者実験計画法の基礎

さて，単一被験者実験計画法といっても，様々な方法

が開発されている。Kearns(1986)は，処遇効果の評価方略を以下のように4つに分け，その各々の場合に利用可能な単一被験者実験計画法を明確にし，それぞれの単一被験者実験計画法の基本的な考え方を示した。その分類をもとに，それぞれの実験計画法の特徴と利用上の留意事項などについてみてみよう。

(1) 処遇－非処遇比較 (treatment-no treatment comparison)

これは，単一被験者実験計画法の最も基本的な方略であり，処遇を行うことは，それを行わないときに比べて，行動の改善をもたらすかどうかを明らかにしようとするときに実施する。非処遇のベースライン期と実験操作や治療処置を行う処遇期とを比較して，その処遇の効果を評価しようとするものである。事前に独立変数としての処遇の与え方と，従属変数としての変容すべき標的行動を明確に規定しておく必要がある。

次の4つの実験計画法が適宜利用できる。

① 除去・反転計画法 (withdrawal and reversal design) …… 基本的には，処遇を除去したあとで，治療効果が逆転するかどうかをみる。これはさらに，1) A－B－A 計画法 (Walker & Buckley, 1968)，2) A－B－A－B計画法 (Hall, et al., 1971)，3) B－A－B計画法(Ayllon & Azrin, 1965) などに分かれる。

A－B－A 計画法では，処遇をまったく行わないベースライン期(A)で行動が安定したことを確認した後，処遇期(B)で治療変数を導入しておいて，次いでそれを除去してベースライン期の状態にもどす(A)ことによって，処遇の導入と除去の効果を分析する。しかしこの実験計画法は，臨床的には，治療が除去された状態で終了するという欠点を持つ。そのため，まず先行して治療処置を開始しておき(B)，次きにその介入を除去し(A)，さらに再び治療変数を導入する(B)というB－A－B計画法や，2回目のベースライン期のあとで，もう一度処遇期を設けるA－B－A－B計画法が開発された。これらの実験計画法によれば，もし必要であれば，さらに治療介入を継続することができるとともに，A－B－A－B計画法ではさらに，2回にわたって，処遇期と非処遇期における標的行動の変化が確認できるという利点がある。

このA－B－A－B計画法には，数多くの変法があることが知られている(例　A－B－C－B計画法：Porterfield, Blunden & Blewitt, 1980；　A－B－A－B′－B″ 計画法：Hopkins, Schutte, & Garton, 1971 など)が，いずれにしても除去計画法では，処遇のあとのベースライン期で処遇が除去されたならば，標的行動はつねにベースラインの水準に戻ることによって，実験条件や治療変数の効果が確かめられるのである。

そのような，標的行動の最初のベースライン水準への反転が起こりにくいときには，反転計画法や多層ベースライン計画法を採用するとよい。前者は，2回目のベースライン期において，標的行動以外のそれと相入れない行動に独立変数を適用する(処遇を実行する)ことによってなされる (Leitenberg, 1973)。

② 多層ベースライン計画法(multiple baseline design) …… 前述のように，ベースライン期で行動の反転が生起しなかったり，しにくかったり，治療介入を停止することが倫理的に望ましくない場合，あるいは，A－B－A－B計画法では時間がかかりすぎるようなときに，この多層ベースライン計画法は用いられる(Bornstein, Bellack, & Hersen, 1977)。そして，機能的に独立した標的行動をとりあげることが可能だったり，均質な被験者が何人かいるようなときに利用できる。

この計画法には，基本的に次の3つのタイプがある。1)行動間多層ベースライン計画法：1人の被験者の各々独立した複数の標的行動に，同一の実験条件や治療変数を順次導入する。2)被験者間多層ベースライン計画法：同一の環境状況のもとで，何人かの均質な被験者の同一の標的行動に対して，特定の実験条件や治療変数を順次導入する。3)事態間多層ベースライン計画法：各々独立した複数の事態において，1人の被験者に特定の実験条件や治療変数を順次導入する。

③ 多層プローブ法(multiple probe technique) …… この手法は，機能的に独立な標的行動を選択することはできるものの，その行動を反復測定すること自体が，その行動の生起に影響を与えることが予想されるような場合に利用する (Horner & Baer, 1978)。ベースライン期の測定を連続することが，その被験者の標的行動の出現を歪曲してしまうおそれがあれば，例えば，3つの行動間に多層プローブ法を用いようとするとき，ベースライン期でのそれぞれの標的行動の測定を，6回，9回，12回と連続して行う代わりに，2回，3回，4回のプローブを3～4回目の測定ごとに実施すればよい。多層プローブ法は，多層ベースライン計画法と同様に，以上のように，行動間はもとより，被験者間，事態間においても適用できる。

また，実験期や治療期での反復測定が，その処遇の効果に影響を及ぼすことが考えられるのであれば，そこでも同じようにプローブを導入すべきであろう。

(2) 成分評価 (component assessment)

複数の実験条件や治療要因を併せた処遇パッケージにおいて，一つひとつの要素が行動変容や治療効果にどの程度の影響を及ぼしているか，そしてまた，処遇パッケージに新たに要素を付け加えることが，処遇の効果を促進するかどうかを検証しようとする場合に利用する。実験や治療の1つあるいはそれ以上の要素の効果を，全体の処遇パッケージとの関係で評価しようとするものである。その際に，それらの要素が処遇パッケージとの組み合わせにおいても，単独ででも検証することができ，かつ被験者間でも繰り返し確認することができるかどうかが，成分評価を行っていくうえで重要な要件となる (Leitenberg, et al., 1968)。次の2つの計画法がある。

① 交互作用計画法(削減) (interaction design, reduction) …… これは，要素の削減の効果をみようとするものである。この実験計画法は，基本的には前に述べたA-B-A計画法と同じ形式を持つものであるが，ベースライン期と処遇期を比較するのではなく，処遇の1つの要素(B)と複数の要素を併せた処遇パッケージ(BC)との比較を行う。例えば，BC-B-BC-Bという実験計画によって，BCの効果とBの処遇の効果との比較が可能となる。

② 交互作用計画法(付加) (interaction design, additive) …… 別の要素を新たに付加したときの効果の評価に使う。例えば，B-BC-B-BCという実験計画において，BとBCの効果を比べることによって，BCがBを上回っていれば，BとCとの間には交互作用があることが示唆される。また例えば，A-B-BC-B-BC-Aのようにベースライン期(A)を設定することで，そのベースライン期とそれぞれの処遇期との比較も可能となる。

だが，実験条件や治療変数の交互作用の効果を分析するにあたって，関連する要因を厳密に統制して，単一被験者実験計画法を用いようとすると，1度に1つ以上の変数を操作してはならないという原則のために，相当の長期間が必要となり，現実的ではなくなってしまうことがある。

(3) 処遇－処遇比較(treatment-treatment comparison)

これは，1人の被験者内での2つあるいはそれ以上の処遇の効果を比較するための実験計画法である。次の2つに分けられる。

① 処遇交替計画法 (alternating treatment design) …… 2つあるいはそれ以上の実験条件や治療法に関して，それぞれの相対的な効果を知りたいときに利用できる (Barlow & Hayes, 1979)。その場合に，各被験者について，実験条件のそれぞれあるいは治療法の各々が迅速に交替できることが条件である。

実験の要因どうしの比較や治療法の間の有効性に関する分析は，従来は集団比較実験計画法を用いてほとんどが行われてきたといってよい。しかし前述のように，統計的推測を中心とする集団比較実験計画法の場合には，個々の集団の平均値を代表値として結果を処理するために，個人のばらつきは誤差として扱われてしまい，得られた結果を一般化するうえで問題があることが示された。

ところがこの処遇交替計画法によれば，1人の被験者に，例えば2種類の異なる治療法を時期をずらして適用することにより，その2つの治療法のその被験者に対する効果が直接に確認でき，またそこでの結果を別の被験者で確かめることを繰り返すことによって，V.で考察するような内的妥当性，外的妥当性ともに高い結果がもたらされることになる。さらに，処遇交替計画法の利点としては，a.治療法どうしを比較したい場合には，つねにどちらかの治療が行われていることになり，ベースライン期を含む計画法のように，治療が除去されることがない(ただし，治療期とベースライン期との交替も可能であるが，その場合にもベースライン期はすぐに治療期と交替する)，b.また，除去計画法に比べて，時間的に早くそれぞれの治療法の効果についての情報が得られる，などがあげられる。

とはいえ，処遇交替計画法を利用するときには，いくつかの点に留意する必要がある。この計画法は，実験の経過や治療セッションの進行に伴う行動の変容・改善に焦点を当てているわけではなく，あくまで2つあるいはそれ以上の実験条件間や治療法間の比較をすることが目的である。ところが，こうした複数の要因を交替に導入する際には，それらの要因が相互に干渉しあって，本来の効果とは別の効果が現れることがある。持ち越し効果 (carryover effects) や順序効果 (order effects) などがそれである (Barlow & Hayes, 1979)。前者は，直前に実施された処遇がその直後の処遇に作用すること，後者は，ある処遇の効果は，それを含めて処遇を実施していくなかでのそれ以前に実施された処遇の効果をも反映しているということである。こうした点を実験的に統制するためには，a.要因を導入する順序は十分にカウンターバランスをとること，b.治療セッションの時間間隔は十分に設けること，などに配慮すべきであろう。

(4) 連続的水準分析 (successive level analysis)

ここに属する評価方略は，1人の被験者の中で，処遇

の水準を連続的に変化させたときの，相対的な効果を検証するためのものである。以下の2つの方法がある。

① 多層プローブ法(multiple probe technique) …… 変化系列における連続する段階の獲得が，処遇によってもたらされるかどうかを確認したいときに利用できる(Horner & Baer, 1978)。この手法を用いるには，処遇の系列におけるそれぞれの段階が独立していること，以前の段階はそれ以後の段階を獲得するための前提となっていること，が必要である。

② 基準変更計画法(changing criterion design) …… 用いている処遇が，次第に獲得しつつある1つの標的行動を，効果的に変容させていくかどうかを調べようとするものである(Deitz & Repp, 1973)。ここでは，標的行動の変化は，基準の水準の変化と一致していること，標的行動は基準のより高い水準でしだいに安定すること，が要求される。

基準変更計画法も，基本的にはA-B-A計画法の考え方がその根底にあり，いわば直前の段階がつねにその直後の段階のベースラインになっているともいえる。そのような直前の段階の基準と比較して，その次の段階ではより高い基準が設定され，その基準で標的行動が安定するまで処遇が継続される。こうして基準をわずかずつ漸次変化させることに応じて，標的行動にその基準と同じ方向での変化が見出せれば，導入したその処遇の効果が確認されたことになる。

Ⅲ. 単一被験者実験計画法の手続きの検討

導入された実験条件や治療変数の有効性の評価が正当になされるためには，採用すべき単一被験者実験計画法において決められている手続きが，正確に実行されなければならない。以下にそれらの手続きのいくつかの部分について検討を加えてみよう。

(1) 標的行動の規定

標的行動とは，実験や治療の介入によって変容することが期待される行動，つまり従属変数のことであり，処遇効果を判定するための標的となる行動である。したがって通常は，観察可能でかつ測定可能な行動が標的行動として選択され，同時にその行動は一義的でしかも理解が容易なことばを使って明確に規定され，その範囲も厳密に示されなければならない。そのことが，信頼性の高い測定値を得るためのまず第一歩となるのである。

(2) 行動観察に関する要因

標的行動の客観的な定義に次いで，その行動を観察・記録する観察者の要因と，観察方法の要因が測定値の信頼性に影響を及ぼすことになる。

① 観察者 …… ふつう，観察者は標的行動の観察・記録の仕方について，事前に訓練を受ける。また実験や処遇の期間中にも，観察者の解釈や主観的判断が混入しないように，再訓練を行うこともある。観察者の被験者への期待や偏見も観察の結果を歪めてしまう原因となる。このような，観察者に由来する観察の正確さを危うくするような要因の影響を取り除くために，観察資料の収集にあたっては，2人以上の観察者に同一行動を観察・記録させ，観察者間の記録の一致度をもとに，観察資料の信頼性の指標にすることが多く行われている(Berk, 1979)。

② 観察方法 …… 行動観察を実施する場所とか，観察する行動の次元(生起頻度，持続時間，潜時など)とか，観察方略(タイムサンプリング法，事象記録法など)とかを決定する必要がある。

(3) ベースラインの安定性

ベースライン期は，それ以後の処遇効果を判定するための基準となる測定値を収集する期間であって，そこで，なんら操作を加えない自然な状態での標的行動の生起頻度などを測定する。そして，その測定値には，基準としての安定性が要求される。処遇の介入は，原則としてベースラインの測定値がはっきりと安定するまで行わない方がよいが，それには例えば治療の待機という倫理的な制約も含まれてくることになるので注意を要する。行動のばらつきが長期間みられたり，大きな変動が続いたり，あるいはベースライン期ですでに行動の改善が示されるようなときには，統計的方法を適用して，処遇期との比較において処遇の効果の有意性を検定することができる(Ⅳを参照)。

(4) 処遇手続きと測定の反復

実験条件や治療変数のような独立変数を操作することを処遇という。

Ⅱでは処遇の評価方略を4分類して，そのそれぞれに対応する単一被験者実験計画法を示した。それによると，処遇の効果を調べるやり方は，a.ベースラインと処遇とを比較する方法，b.処遇どうしを比較する方法，に大別できよう。前者には除去計画法，反転計画法などがあり，後者には処遇交替計画法などがあった。

除去計画法の手続きは，導入されていた条件や変数を2回目のベースライン期以降で，単純に取り除くことである。それによって，標的行動の生起頻度が最初のベースラインの水準に戻ればよい。しかし，処遇を除去することが被験者に不都合だったり(問題行動の再発など)，

行動が学習されてしまってベースラインの水準まで戻らないような場合には，複数のベースライン期を設けるとか，処遇を交替させるなどの手続きを採用する。

多層ベースライン計画法の手続きでは，ベースライン期の長さを少しずつずらしていき，処遇を開始した時点に合致して，それぞれの標的行動に変容が見られ始めればよい。

なお処遇期の長さに関しては，特にきまりはないが，めやすとしては測定値が安定するまで継続すべきであろう。

Ⅳ. 処遇効果の統計による検定

これまで，単一被験者実験計画法によって処遇効果を判定しようとするときには，統計的な検定は，ほとんど用いられてこなかったといってよい。その理由のひとつは，この実験計画法を利用しようとする研究者や臨床家のほとんどは，一人ひとりの個人における行動の記述やその変容に関心があり，個人差を誤差として処理してしまい，平均値で行動を説明しようとする統計的方法に少なからぬ低抗があったこと (Michael, 1974)，そしてまた，1人もしくはごく少数の被験者を対象とする場合には，推測統計の手法は適用できないとする誤解があったためであろうとされている (Gentile, Roden, & Klein, 1972)。

(1) 効果の視覚的分析

処遇の効果は，今日でも一般的には，測定値をグラフにおとし，その変化をもっぱら目で見て判断するという視覚的分析に頼っている。もちろんこの分析によっても，目で見て認める一定の見方さえ理解していれば，原則として，例えば，ある治療法を用いた介入の効果は具体的に把握できよう。

しかしながら，場合によっては，グラフを見ただけでは処遇の効果の判定が困難なことがある。測定値の変動が非常に大きいとき，あるいは測定値がすでに改善傾向を示すベースライン期に続いて，処遇期が導入されているときなどがそれである。

(2) 統計的分析の適用

こうしたときに，単一被験者実験計画法においても，推測統計による有意性検定を適用することがある。こうして，視覚的分析と統計的検定とは決して背反するものではなく，むしろ相互に補完的に用いていくべきである。統計的検定によれば，採用する統計的手法さえ決定してしまえば，その分析の過程に個人の判断が混在する余地はなくなり，同一の結果がもたらされることになる。

次のような統計的手法を含めて，統計的分析が実行される。

① t検定 …… A-B計画法で，ベースライン期(A)における測定値の平均と，処遇期(B)における測定値の平均の差の検定を行う。それらの平均値の差が統計的に有意であるかどうかを検定するわけである。

② 分散分析 …… 例えばA-B-A-B計画法で，それぞれの手続き期を比較するのに，分散分析を実施する。

以上①，②のような適用は，しかしながら，単一被験者実験計画法による測定値の収集は，1人の被験者において反復測定されたものであるため，測定値が相互に独立であるとはいえず，系列依存性をもつという困難な点がある。この系列依存性が有意でない場合にのみ分散分析は利用すべきだとの主張 (Hartmann, 1974) もあるが，A-B-A-B計画法への分散分析の適用については，本来分散分析の前提条件を満たしていないので用いるべきではない，という指摘もある (Michael, 1974)。

③ 時系列分析 …… 時系列分析 (time-series analysis) というのは，1人の被験者のそれぞれの処遇期の測定値を，時間の経過につれて分析することである (Gottman, 1980)。この分析法は，系列依存性を処理することができるという大きな利点がある反面，多くの測定値の数（それぞれの処遇期において，少なくとも50以上）を必要とするという困難な点ももっている。

④ C統計 …… ③の時系列分析では，数多くの測定値が必要であった。だがこのC統計による分析では，測定値の数は8個以上あれば適用できるとされている (Tryon, 1982 ; 河合他, 1988)。しかし一方でこのC統計は，有意性の検出が不安定であることが指摘されている (Blumberg, 1984)。

以上，単一被験者実験計画法で用いられる統計的検定のいくつかについて述べた。しかしここでも再度，有意性が得られたときのその有意性の意味について確認してみると，それは，a.実験条件や治療変数の違いによって，統計的に有意な変化がもたらされたことだけを表すものであり，その行動変化の原因については何も述べていない。b.行動変化が統計的に有意であるということだけを示し，その変化の臨床的な意義については何も述べていない，ということである。そうした行動変化に関する因果分析や臨床的意義については，実験計画法での実験・治療変数の操作をとおして，また日常生活場面への適応の状態などを観察することによって，検討されなければならない。

V. 内的妥当性と外的妥当性

実験条件が及ぼす影響あるいは処遇の効果，さらにはその研究自体の成果は，得られた測定値の信頼性と妥当性に深く係わりをもつものである。信頼性と妥当性ともに高い情報をもとに研究が展開されるのでなければ，そこでの問題の本質的な解明はありえない（成田，1986）。

(1) 測定値の信頼性

測定された値が測定誤差を含まないほど，その測定値は真の値に近く，信頼性が高いという。標的行動の観察にあたっての測定値の信頼性に係わる要因としては，a.標的行動の規定の仕方，b.観察・測定に使う用具やその方略，c.観察者の観察の熟達度，などがあげられる。これらの要因については，Ⅲ.でも述べたように，とにかくまずは，十分に信頼性の高い測定値が得られるように，条件設定をしていく必要がある。そしてこの信頼のできる測定値を収集することが，次に述べるような，内的妥当性を高める前提の要件となっているのである。ところが，妥当性とは測定しようとしたことをどの程度まで測定できたかに関する概念であるが，測定値の信頼性が高いことは，必ずしも妥当性が高いことを保証するものではない。

以下では，集団比較実験計画法と単一被験者実験計画法とにおける内的妥当性および外的妥当性について，比較検討してみよう。

(2) 内的妥当性

内的妥当性とは，実験や治療において，直接に操作した実験条件や治療変数，つまり独立変数だけによって，もたらされた結果がどれだけ説明できるかということである。

集団比較実験計画法では，直接操作する独立変数の他にも，結果に影響を及ぼすおそれのある要因をも含めて，それぞれの群に被験者を無作為に割り付けることによって，それらの要因を統制しようとする。そして，各々の群間での従属変数の変化に統計的な有意性検定を適用するわけであるが，実はここで得られる情報は，先に述べたように，独立変数の効果を直接に説明するものではない。

それに対して，単一被験者実験計画法においては，例えばA-B-A-B計画法の場合にみるように，処遇期(B)で変数を直接操作し，その際の従属変数の変化を，変数の操作をまったく行わないベースライン期(A)での従属変数の変化とを比較することを，2回にわたって繰り返すことによって，処遇期で導入した変数操作の影響を直接に指摘することができる。このように，同じ1人の被験者のなかで変数の操作を反復することで，処遇の効果をじかに確認できるとともに，さらに別の被験者でも同様に，その処遇の効果が確認されれば，その操作した独立変数が，従属変数である標的行動の変容に影響を与えるだろう確かさは，一層強固なものとなる。

(3) 外的妥当性

外的妥当性とは，ある特定の実験や治療でもたらされた結果が，そこで用いられた被験者，治療者，実験事態，治療場面などの違いを越えて一般化できる程度を示す。内的妥当性が高いことが，外的妥当性を高めるための重要な要件である。

単一被験者実験計画法では，1人だけの被験者を対象に実験操作を行うために，他の被験者や環境へと一般化することは困難があるとの印象を持つかもしれない。しかし実際には，ある1人の被験者で得られた結果を，別の被験者に適用しその結果を確認する，そしてさらに別の被験者で確認するというように，被験者の数を増やしていくことで，外的妥当性は高められていくのである。

他方，集団比較実験計画法においては，結果を一般化するためには，それに関連する要因を一つひとつ操作し，実験条件を設定しなおして，無数の集団比較実験を実施する必要が出てくる。しかも，標本の単純無作為抽出の困難さのゆえに，集団比較実験をたとえ何度も繰り返しても，一般化の程度が増すわけではない。また，例え標本数を増やしても，そのことだけでは，その標本が属していた母集団についてだけ，一般化の程度が増すのみである。さらに，測定値は平均値で示されることとなり，一人ひとりの相違は誤差として無視されてしまう。従って，そこでの結果は個別に適用できなくなる。また群間に統計的な有意性が得られたからといって，そのことは，一人ひとりの被験者にそれと同様な違いがあることを示すものではない。

これに対して単一被験者実験計画法では，ある被験者から得られた結果は，その被験者にとっての事態でいつも適用できるし，変数操作を反復させたり，その変数操作を部分的に変更させることなどによって，結果の一般化の程度は益々増大することになる。

また，もう一方では，実験操作や治療によって変容された行動の般化や維持を，単一被験者において評価するための方略の検討もなされている(Barrios & Hartmann, 1988)。

おわりに－行動の変容と臨床的意義

本稿では，集団比較実験計画法が内在させている難点

を取り上げながら，単一被験者実験計画法を利用するうえで留意すべきいくつかの問題について述べた。

集団比較実験計画法は，再三述べてきたように，各群の平均値や分散をもとに，統計的有意性の検定を行い，その被験者が属していた母集団の特徴を推測するために利用する。そこでは，測定値は平均にすべて丸められてしまい，たとえ，群の大勢とは逆の方向の行動変容が見られたとしても，また個人の特徴も，誤差として扱われてしまうのである。

それゆえに，統計的な有意性検定からは，各々の個人に直接適用できる結果は得られない。実は，一人ひとりの特性を把握することを主眼とする臨床的接近のためには，統計的な有意性はほとんど役にたたないといえる。それにもかかわらず，統計的に有意性が見出されると，その情報がそのまま臨床に適応できるかのように，誤って判断されてしまうことがある。統計的有意性の示すなかには，そこでの処遇や介入が臨床的な問題の解決に有効性を持つという意味はない。

臨床家は，他方，いつも被験者の問題行動を前にして，その解決のための，直接の具体的な手段を希求しているといえる。個体の特徴を独自のものとして捉え，一人ひとりの要求に則した臨床的対応を実行すること，それが臨床的意義である。

こうした臨床的に有用な情報の収集のためには，単一被験者実験計画法は，大いに利便があるといえよう。

引用文献

Ayllon, T., & Azrin, N.H. 1965 The measurement and reinforcement of behavior of psychotics. Journal of Experimental Analysis of Behavior, 8, 357-383.

Baer, D.M., Wolf, M.M., & Risley, T.R. 1968 Some current dimensions of applied behavior analysis. Journal of Applied Behavior Analysis, 1, 91-97.

Barlow, D.H., & Hayes, S.C. 1979 Alternating treatment design: One strategy for comparing the effects of two treatments in a single subject. Journal of Applied Behavior Analysis, 12, 199-210.

Barlow, D.H., & Hersen, M. 1984 Single case experimental designs: Strategies for studying behavior change (2nd ed.). New York: Pergamon Press.

Barrios, B.A., & Hartmann, D.P. 1988 Recent developments in single-subject methodology: Methods for analysing generalization, maintenance, and multicomponent treatments. In M. Hersen, R.M. Eisler, & P.M. Miller (Eds.), Progress in behavior modification. vol. 22, pp. 11-47, New York: Academic Press.

Berk, R.A. 1979 Generalizability of behavioral observations: A classification of interobserver agreement and interobserver reliability. American Journal of Mental Deficiency, 83, 460-472.

Blumberg, C.J. 1984 Comments on "A simplified time-series analysis for evaluating treatment interventions." Journal of Applied Behavior Analysis, 17, 539-542.

Bornstein, M.R., Bellack, A.S., & Hersen, M. 1977 Social-skills training for unassertive children: A multi-baseline analysis. Journal of Applied Behavior Analysis, 10, 183-195.

Deitz, S.M., & Repp, A.C. 1973 Decreasing classroom misbehavior through the use of DRL schedules of reinforcement. Journal of Applied Behavior Analysis, 6, 457-463.

Fisher, R.A. 1925 On the mathematical foundations of the theory of statistics. In Cambridge Philosophical Society (Ed.), Theory of statistical estimation. Proceedings of the Cambridge Philosophical Society, England.

Gentile, J.R., Roden, A.H., & Klein, R.D. 1972 An analysis of variance model for the intrasubject replication design. Journal of Applied Behavior Analysis, 5, 193-198.

Gottman, J.M. 1980 Time-series analysis: A comprehensive introduction for social scientists. Cambridge: Cambridge University Press.

Hall, R.V., Fox, R., Willard, D., Goldmith, L., Emerson, M., Owen, M., Davis, F., & Porsia, E. 1971 The teacher as observer and experimenter in the modification of disputing and talking-out behavior. Journal of Applied Behavior Analysis, 4, 141-149.

Hartmann, D.P. 1974 Forcing square pegs into round holes: Some comments of "An analysis-of-variance model for the intrasubject replication design". Journal of Applied Behavior Analysis, 7, 635-638.

Hopkins, B.L., Schutte, R.C., & Garton, K.L. 1971 The effects of access to a playroom on the rate and quality of printing and writing of first-and-second grade students. Journal of Applied Behavior Analysis, 4, 77-87.

Horner, R.D., & Baer, D.M. 1978 Multiple-probe

technique: A variation of the multiple baseline. Journal of Applied Behavior Analysis, 11, 189-196.

河合伊六・河本　肇・大河内浩人　1988　単一事例計画法における処遇効果のC統計による検定　行動分析学研究, 2, 36-47.

Kearns, K.P. 1986 Flexibility of single-subject experimental designs. part II: Design selection and arrangement of experimental phase. Journal of Speech and Hearing Disorders, 51, 204-214.

Kratochwill, T.R. 1978 Single subject research: Strategies for evaluating change. New York: Academic Press.

Leitenberg, H. 1973 The use of single-case methodology in psychotherapy research. Journal of Abnormal Psychology, 82, 87-101.

Leitenberg, H., Agras, W.S., Thomson, L.E., & Wright, D.E. 1968 Feedback in behavior modification. An experimental analysis of two phobic cases. Journal of Applied Behavior Analysis, 1, 131-137.

Michael, J. 1974 Statistical inference for individual organism research: Some reactions to a suggestion by Gentile, Roden, and Klein. Journal of Applied Behavior Analysis, 7, 627-628.

Morrison, D.E., & Henkel, R.E. 1970 The significant test controversy. Chicago: Aldine.

成田　益　1986　単一被験者実験計画法における内的妥当性と外的妥当性に関する一考察　特殊教育学研究, 23(4), 37-43.

Porterfield, J., Blunden, R., & Blewitt, E. 1980 Improving environments for profoundly handicapped adulsts: Using prompts and social attention to maintain high group engagement. Behavior Modification, 4, 225-241.

Sidman, M. 1960 Tactics of scientific research: Evaluating experimental data in psychology. New York: Basic Books.

Skinner, B.F. 1953 Science and human behavior. New York: Macmillan.

橘　敏明　1986　医学・教育学・心理学にみられる統計的検定の誤用と弊害　医歯薬出版社

Tryon, W.W. 1982 A simplified time-series analysis for evaluating treatment intervention. Journal of Applied Bahavior Analysis, 15, 423-429.

Walker, H.M., & Buckley, N.K. 1968 The use of positive reinforcement in conditioning attending behavior. Journal of Applied Behavior Analysis, 1, 245-250.

APPLICATION OF SINGLE SUBJECT EXPERIMENTAL DESIGNS AND ITS ISSUES

Naoji SHIMIZU
Division for the Education of the Mentally Retarded

ABSTRACT

The present article focuses the issues related to single subject experimental designs in which experiments are performed on individual subjects.

Although, in the area of research in psychology and special education, group comparison experimental designs are more often used than single subject experimental designs, it is recognized that the application of those group designs to applied problems pose many difficulties. On a more practical level, the collection of large numbers of subjects homogeneous for a particular experimental condition is often very difficult, and the averaging of individual subjects and the use of statistical significance tests also bring up some related problems.

In the development of behavior analysis since the 1960's, new experimental design alternatives have appeared which emphasize direct measurement of treatment effects and causal analysis between independent variables and dependent variables.

In describing the methodology, general issues in single subject experimental designs were discussed.

Key words: single subject experimental design, group comparison experimental design, statistical significance test, treatment effects, reliability, validity